IN THE CLAIMS

1. (Currently amended) A method of optimizing data mining in a computer, the data mining being performed by the computer to detect one or more outliers within a high dimensional data set stored on a data storage device coupled to the computer, the data set representing a population of ~~individuals~~ persons and the one or more outliers representing one or more ~~individuals~~ persons within the population of ~~individuals~~ persons, the method comprising the steps of:

determining one or more subsets of dimensions and corresponding ranges in the data set which are sparse in density using an algorithm comprising at least one of the processes of solution recombination, selection and mutation over a population of multiple solutions; and

determining one or more data points in the data set which contain these subsets of dimensions and corresponding ranges, the one or more data points being identified as the one or more outliers;

wherein the sets of dimensions and corresponding ranges in which the data is sparse in density is quantified by a sparsity coefficient measure.

2. (Original) The method of claim 1, wherein a range is defined as a set of contiguous values on a given dimension.

3. (Canceled)

4. (Previously presented) The method of claim 1, wherein the sparsity coefficient measure $S(D)$ is defined as $\frac{n(D) - N * f^k}{\sqrt{N * f^k * (1 - f^k)}}$, where k represents the number of dimensions in the data set, f represents the fraction of data points in each range, N is the total number of data points in the data set, and $n(D)$ is the number of data points in a set of dimensions D.

5. (Previously presented) The method of claim 1, wherein a given sparsity coefficient measure is inversely proportional to the number of data points in a given set of dimensions and corresponding ranges.

6. (Original) The method of claim 1, wherein a set of dimensions is determined using an algorithm which uses the processes of solution recombination, selection and mutation over a population of multiple solutions.

7. (Original) The method of claim 6, wherein the process of solution recombination comprises combining characteristics of two solutions in order to create two new solutions.

8. (Original) The method of claim 6, wherein the process of mutation comprises changing a particular characteristic of a solution in order to result in a new solution.

9. (Original) The method of claim 6, wherein the process of selection comprises biasing the population in order to favor solutions which are more optimum.

10. (Currently amended) A method of optimizing data mining in a computer, the data mining being performed by the computer to detect one or more outliers within a high dimensional data set stored on a data storage device coupled to the computer, the data set representing a population of ~~individuals~~ persons and the one or more outliers representing one or more ~~individuals~~ persons within the population of ~~individuals~~ persons, the method comprising the steps of:

identifying and mining one or more sub-patterns in the data set which have abnormally low presence not due to randomness using an algorithm comprising at least one of the processes of solution recombination, selection and mutation over a population of multiple solutions; and

identifying one or more records which have the one or more sub-patterns present in them as the one or more outliers;

wherein the abnormally low presence is quantified by a sparsity coefficient measure.

11. (Currently amended) Apparatus for optimizing data mining to detect one or more outliers within a high dimensional data set, the data set representing a population of ~~individuals~~ persons and the one or more outliers representing one or more ~~individuals~~ persons within the population of ~~individuals~~ persons, comprising:

a computer having a memory and a data storage device coupled thereto, wherein the data storage device stores the data set; and

one or more computer programs, performed by the computer, for: (i) determining one or more subsets of dimensions and corresponding ranges in the data set which are sparse in density using an algorithm comprising at least one of the processes of solution recombination, selection and mutation over a population of multiple solutions; and (ii) determining one or more data points in the data set which contain these subsets of dimensions and corresponding ranges, the one or more data points being identified as representing the one or more the one or more outliers;

wherein the sets of dimensions and corresponding ranges in which the data is sparse in density is quantified by a sparsity coefficient measure.

12. (Previously presented) The apparatus of claim 11, wherein a range is defined as a set of contiguous values on a given dimension.

13. (Canceled)

14. (Previously presented) The apparatus of claim 11, wherein the sparsity coefficient measure $S(D)$ is defined as $\frac{n(D) - N * f^k}{\sqrt{N * f^k * (1 - f^k)}}$, where k represents the number of dimensions in the data set, f represents the fraction of data points in each range, N is the total number of data points in the data set, and $n(D)$ is the number of data points in a set of dimensions D.

15. (Previously presented) The apparatus of claim 11, wherein a given sparsity coefficient measure is inversely proportional to the number of data points in a given set of dimensions and corresponding ranges.

16. (Original) The apparatus of claim 11, wherein a set of dimensions is determined using an algorithm which uses the processes of solution recombination, selection and mutation over a population of multiple solutions.

17. (Original) The apparatus of claim 16, wherein the process of solution recombination comprises combining characteristics of two solutions in order to create two new solutions.

18. (Original) The apparatus of claim 16, wherein the process of mutation comprises changing a particular characteristic of a solution in order to result in a new solution.

19. (Original) The apparatus of claim 16, wherein the process of selection comprises biasing the population in order to favor solutions which are more optimum.

20. (Currently amended) Apparatus for optimizing data mining to detect one or more outliers within a high dimensional data set, the data set representing a population of ~~individuals~~ persons and the one or more outliers representing one or more ~~individuals~~ persons within the population of ~~individuals~~ persons, comprising:

a computer having a memory and a data storage device coupled thereto, wherein the data storage device stores; and

one or more computer programs, performed by the computer for: (i) identifying and mining one or more sub-patterns in the data set which have abnormally low presence not due to randomness using an algorithm comprising at least one of the processes of solution recombination, selection and mutation over a population of multiple solutions; and (ii) identifying one or more records which have the one or more sub-patterns present in them as the one or more outliers;

wherein the abnormally low presence is quantified by a sparsity coefficient measure.

21. (Currently amended) An article of manufacture comprising a program storage medium readable by a computer and embodying one or more instructions executable by the computer to perform method steps for optimizing data mining, the data mining being performed by the computer to detect one or more outliers within a high dimensional data set stored on a data storage device coupled to the computer, the data set representing a population of ~~individuals~~ persons and the one or more outliers representing one or more ~~individuals~~ persons within the population of ~~individuals~~ persons, the method comprising the steps of:

determining one or more subsets of dimensions and corresponding ranges in the data set which are sparse in density using an algorithm comprising at least one of the processes of solution recombination, selection and mutation over a population of multiple solutions; and

determining one or more data points in the data set which contain these subsets of dimensions and corresponding ranges, the one or more data points being identified as the one or more outliers;

wherein the sets of dimensions and corresponding ranges in which the data is sparse in density is quantified by a sparsity coefficient measure.

22. (Original) The article of claim 21, wherein a range is defined as a set of contiguous values on a given dimension.

23. (Canceled)

24. (Previously presented) The article of claim 21, wherein the sparsity coefficient measure $S(D)$ is defined as $\frac{n(D) - N * f^k}{\sqrt{N * f^k * (1 - f^k)}}$, where k represents the number of dimensions in the data set, f represents the fraction of data points in each range, N is the

total number of data points in the data set, and $n(D)$ is the number of data points in a set of dimensions D.

25. (Previously presented) The article of claim 21, wherein a given sparsity coefficient measure is inversely proportional to the number of data points in a given set of dimensions and corresponding ranges.

26. (Original) The article of claim 21, wherein a set of dimensions is determined using an algorithm which uses the processes of solution recombination, selection and mutation over a population of multiple solutions.

27. (Original) The article of claim 26, wherein the process of solution recombination comprises combining characteristics of two solutions in order to create two new solutions.

28. (Original) The article of claim 26, wherein the process of mutation comprises changing a particular characteristic of a solution in order to result in a new solution.

29. (Original) The article of claim 26, wherein the process of selection comprises biasing the population in order to favor solutions which are more optimum.

30. (Currently amended) An article of manufacture comprising a program storage medium readable by a computer and embodying one or more instructions executable by the computer to perform method steps for optimizing data mining, the data mining being performed by the computer to detect one or more outliers within a high dimensional data set stored on a data storage device coupled to the computer, the data set representing a population of ~~individuals~~ persons and the one or more outliers representing one or more ~~individuals~~ persons within the population of ~~individuals~~ persons, the method comprising the steps of:

identifying and mining one or more sub-patterns in the data set which have abnormally low presence not due to randomness using an algorithm comprising at least

one of the processes of solution recombination, selection and mutation over a population of multiple solutions; and

identifying one or more records which have the one or more sub-patterns present in them as the one or more outliers;

wherein the abnormally low presence is quantified by a sparsity coefficient measure.